

UNITED STATES PATENT APPLICATION

of

John A. Tardif

for

SYSTEMS AND METHODS FOR MPEG SUBSAMPLE DECODING

WORKMAN, NYDEGGER & SEELEY

A PROFESSIONAL CORPORATION

ATTORNEYS AT LAW

1000 EAGLE GATE TOWER

60 EAST SOUTH TEMPLE

SALT LAKE CITY, UTAH 84111

BACKGROUND OF THE INVENTION

1. The Field of the Invention

The present invention relates to the field of digital video. More specifically, the present invention relates to reducing the memory bandwidth and memory footprint needed to decode and display MPEG video streams.

2. The Related Art

High Definition Television ("HDTV") is a type of television that provides much better resolution than televisions based on National Television Systems Committee ("NTSC") standard. Although there are a number of competing HDTV standards, all standards support a wider screen than NTSC and roughly twice the resolution. Sending uncompressed analog HDTV data requires a bandwidth of about 18Mhz. However, current terrestrial channel allocations are limited to 6Mhz. As a result, HDTV video frames are digitized and then compressed before they are transmitted and then decompressed when they reach a receiving device, such as an HDTV television.

One widely used compression method is based on the Moving Pictures Experts Group standard and is commonly referred to as MPEG. MPEG employs interframe encoding, which means some of the frames are used as reference frames for other frames in compressed video data. An MPEG video bit stream includes I-frames, P-frames and B-frames. I-frames and P-frames can be used as a reference for other frames, hence they are known collectively as reference frames. I-frames or "Intraframes," are independent frames that may be encoded and decoded independently without referring to any other frames in the MPEG video bit stream. P-frames, or "Predictive" frames, are encoded and decoded using the previous reference frame, be it an I-frame or a P-frame. B-frames, or "Bi-

1 directionally predictive” frames, are reproduced using reference frames that are the closest
2 temporally previous to and/or subsequent to the B-frame. Since I-frames do not reference
3 other frames for information, I-frames are typically substantially larger in size than P-
4 frames and B-frames.

5 An MPEG video frame having the YUV420 format includes regions designated as
6 macroblocks having a size of 16 pixels by 16 lines. Within each macroblock, there are six
7 8x8 blocks of data, four for luminance components, and two for subsampled chrominance
8 data.

9 As shown in Figure 1a, an MPEG encoding system, such as MPEG encoder 100,
10 receives video data 104, which is a sequence of video images. MPEG encoder 100
11 typically includes discrete cosine transform (DCT) module 101, motion vector generation
12 module 102 and a picture type determination module 103, which separate video data 104
13 into different requisite parts. DCT module 101 is used to transform blocks of the video
14 data from the spatial domain into a frequency domain representation of the same blocks.
15 Motion vector generation module 102 is used to generate motion vectors, which represent
16 motion between macroblock regions in the frames of video data 104. Picture type
17 determination module 103 determines which frames should be used as reference frames (I-
18 frames). After being encoded, MPEG video bit stream 105 includes frequency coefficients
19 106, motion vectors 107, and header information 108, which specifies size, picture coding
20 type, etc.

21 To reconstruct the original sequence of video images, inverse operations are
22 performed, as illustrated by MPEG decoder 110 in Figure 1b. Frequency coefficients 106
23 are dequantized and passed through inverse discrete cosine transform (IDCT) module 111,
24 thus converting them back into spatial domain representations. Motion vector module 112

uses header information 108 and motion vectors 107 to recreate the macroblocks of P-frames and B-frames. The outputs from IDCT module 111 and motion vector module 112 are then summed by summer 113 to generate reconstructed output 114. Reconstructed output 114 is a sequence of video images similar to video data 104 from Figure 1a, and can be displayed on a display device.

HDTV video frames consist of 1088 lines, each having 1920 pixels, which results in approximately two million pixels per frame. As alluded to previously, since MPEG uses a YUV420 color space, one pixel is represented using 1.5 bytes. Thus a single HDTV frame uses 3 MB. Since two reference images are maintained in order to correctly decode B-frames, and double buffering is usually desired at the output of the MPEG decoder so that it can decode an image while the video output displays the previous image, this implies that 12 MB of storage are needed for the frames of video data generated by the MPEG decoding process and the associated reference buffers. Equivalently, a standard resolution, NTSC frame consists of 480 lines, each having 720 pixels, or approximately 350,000 pixels per frame. With the YUV420 format, this means that each NTSC frame uses about 520 KB of memory. As a result, the decoder and display device for processing and displaying NTSC video data encoded using MPEG requires about 2.1 MB of storage for the frames of video data generated by the MPEG decoding process and the associated reference buffers.

In North America and other regions, video frames are conventionally sent at the rate of thirty frames per second. The memory bandwidth needed to store the output video data is ~90 MB/sec for HDTV and ~15.5 MB/sec for standard resolution broadcasts. In addition, MPEG decoding requires that predictions be made from reference images.

1 During periods of worst case predictions, up to 4 times that amount of bandwidth may
2 need to be supported (depending on memory subsystem).

3 Due to the large installed base of NTSC televisions, it may often be the case that a
4 video images having been compressed using MPEG and formatted for display on a HDTV
5 device may need to be displayed on a lower resolution NTSC television. It may also be the
6 case that any such video images may need to be displayed in a lower resolution, such as
7 when using picture-in-picture functionality of a television. A conventional method for
8 supporting this application is to fully decode the transmitted images at their native
9 resolution, then resample the transmitted images to the required display resolution.
10 However, decoding an MPEG video bit stream having full HDTV formatting and then
11 resampling to a lower resolution wastes memory resources and computational resources,
12 since the receiving device cannot display the full resolution of the image. As a result,
13 certain methods could be used to reduce the memory footprint, memory throughput and the
14 processing requirements for this application. Figures 1c and 1d illustrate methods that
15 could be used to reduce the memory footprint, memory throughput and the processing
16 requirements. It is noted that the following methods do not necessarily represent prior art
17 with respect to the present invention, but are presented herein to illustrate the advantages
18 of the present invention compared to other approaches that could be implemented.

19 One class of methods involve modifications to the video data before transmission,
20 such resampling to the desired display resolution or hierarchical encoding. All these
21 methods can produce very good image quality. However, they are all limited in that they
22 cannot support any and all desired output resolutions simultaneously in a broadcast
23 environment, since the processing is performed prior to transmission, rather than at the
24

1 decoding or display devices where the images are to be displayed. Also, most of these
2 methods would involve non-standard profiles of MPEG video compression.

3 Another class of methods uses algorithms that are executed on the receiver. These
4 methods attempt to reduce the size of the decompressed video images and the associated
5 reference buffers. These reductions in size have an effect of reducing memory footprint
6 for the buffers, reducing memory bandwidth for processing the decompressed video
7 images, and reducing image resampling computational requirements. Most of these
8 algorithms entail reducing the number of samples in frames in the horizontal and vertical
9 directions by a factor of 2^N , where N is normally 1.

10 One method, as shown in Figure 1c, involves resampling the video frame after the
11 frame has been decompressed using MPEG decoder 110 and prior to storing the
12 decompressed frame in memory. This method can reduce memory footprint by a factor of
13 four if the video frame is subsampled by a factor of two in the horizontal and vertical
14 directions. This involves subsampling motion vectors 107 by a factor of two, then
15 upsampling fetched motion reconstruction data 115 by factor of two in the horizontal and
16 vertical directions. In a parallel operation, frequency coefficients 106 are dequantized and
17 passed through IDCT module 111, which converts the coefficients back into spatial
18 domain data 116. Spatial domain data 116 and the upsampled fetched motion
19 reconstruction data 115 are then summed by summer 113. The output of summer 113 is
20 then subsampled by a factor of two in each direction. This method is hindered by the fact
21 that the output subsampling may require some extra buffering in order to allowing vertical
22 filtering. Also, for relatively static scenes or constant pans, the error terms coming from
23 the IDCT are nearly zero, which results in effectively the same image data being
24 upsampled and downsampled many generations. This generational loss progressively

1 degrades the image quality until an I-frame is decoded, in which case the image is
2 refreshed. This results in a “beating” effect that is most noticeable and irritating to the
3 viewer.

4 Another conventional method, as shown in Figure 1d, involves ignoring all high
5 frequency coefficients and using a 4x4 IDCT instead of a 8x8 IDCT. Similar to the
6 method in Figure 1c, motion vectors 107 are downsampled. However, fetched motion
7 reconstruction data 115 may be directly summed with spatial domain data 116 without
8 requiring post processing of the summed result, which reduces the effect of generational
9 loss described above in reference to Figure 1c. This method reduces memory footprint by
10 a factor of four and significantly reduces the number of computations. However, simply
11 ignoring the high frequency IDCT components can produce some significant artifacts at
12 the boundaries of blocks and macroblocks in the decoded image (otherwise know as
13 “block” artifacts). These artifacts in turn can significantly affect subsequent images that
14 use the previous ones as references. Also, the 4x4 IDCT is slightly different from the 8x8
15 IDCT and for some hardware implementations is not easily changed.

16 In view of the foregoing, there exists a need for systems and methods for efficiently
17 subsampling video data in preparation for displaying the video data on devices of lower
18 resolution than what the video data was originally encoded for, thereby reducing the
19 memory bandwidth and memory footprint needed to process the video data without
20 appreciably reducing the quality of the output.

SUMMARY OF THE INVENTION

The principles of the present invention provide for efficiently subsampling video data for display at a lower resolution than what the video data was originally encoded for. MPEG video bit streams are decoded by dequantizing frequency coefficients and passing them through an IDCT module to convert the coefficients back into the spatial domain representations, passing header information and motion vectors through a motion vector prediction module to fetch predictions from previously decoded images, and then summing the spatial domain representations and the fetched prediction data. The present invention performs decimation on the output of the IDCT module before it is summed with the results of the prediction module. Using this method, predictive and bi-directionally predictive frames do not significantly degrade beyond any initial blurring or softening associated with corresponding reference frames, particularly where the video frames represent relatively static images or constant pans.

Decimation by any factor may be performed in both the horizontal and vertical directions to generate various reductions in the memory bandwidth and footprint needed for image processing. For instance, one-dimensional subsampling by a factor of two, performed in the horizontal direction, yields a two-fold reduction in the memory bandwidth and footprint needed for image processing. Two-dimensional subsampling, or subsampling in both the horizontal and vertical directions, yields a four-fold reduction. Two-dimensional subsampling is most useful in frame-based video data, although it can also be used in field-interlaced video data.

Subsampling the output of the IDCT reduces the need for the IDCT to perform additional computations, yet the reconstructed output maintains much of the image quality associated with reference frames. Additional features and advantages of the invention will

1 be set forth in the description, which follows, and in part will be obvious from the
2 description, or may be learned by the practice of the invention. The features and
3 advantages of the invention may be realized and obtained by means of the instruments and
4 combinations particularly pointed out in the appended claims. These and other features of
5 the present invention will become more fully apparent from the following description and
6 appended claims, or may be learned by the practice of the invention as set forth hereinafter.

BRIEF DESCRIPTION OF THE DRAWINGS

In order that the manner in which the above-recited and other advantages and features of the invention are obtained, a more particular description of the invention briefly described above will be rendered by reference to specific embodiments thereof, which are illustrated, in the appended drawings. Understanding that these drawings depict only typical embodiments of the invention and are not therefore to be considered to be limiting of its scope, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

Figure 1a illustrates a conventional system used to encode an MPEG video bit stream.

Figure 1b illustrates a conventional system used to decode an MPEG video bit stream.

Figure 1c illustrates a system for reducing the memory bandwidth and footprint needed to display video images that have been encoded in an MPEG video bit stream, which subsamples the reconstructed output of an MPEG decoder.

Figure 1d illustrates a system for reducing the memory bandwidth and footprint needed to display video images that have been encoded in an MPEG video bit stream, which ignores some frequency coefficients included in the MPEG video bit stream.

Figure 2 illustrates an exemplary system that provides a suitable operating environment for the present invention.

Figure 3 illustrates a system of the invention that can reduce the memory bandwidth and footprint needed to display video images that have been encoded in an MPEG video bit stream and with an HDTV format on a device of lower resolution.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Figure 4 is a flow diagram illustrating a method for decoding and subsampling the MPEG video bit stream using the system of Figure 3.

Figure 5a illustrates a technique for subsampling video data in the horizontal direction according to one embodiment of the invention.

Figure 5b illustrates a technique for subsampling non-interlaced video data in the vertical direction according to one embodiment of the invention.

Figure 5c illustrates a technique for subsampling interlaced video data in the vertical direction according to one embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention extends to both methods and systems for subsampling and displaying video images that have been encoded using an HDTV format and in an MPEG video bit stream on a device of lower resolution, such as standard television, or on a picture-in-picture display. The embodiments of the present invention may comprise a special purpose or general-purpose processing device or computer including various computer hardware components, as discussed in greater detail below. The embodiments may further comprise multiple computers linked in a networked environment. Set top boxes that enhance the capabilities of conventional televisions represent an example of a special purpose computer.

Embodiments within the scope of the present invention also include computer-readable media for carrying or having computer-executable instructions or data structures stored thereon. Such computer-readable media can be any available media that can be accessed by a general purpose or special purpose computer. By way of example, and not limitation, such computer-readable media can comprise physical storage media such as RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other medium that can be used to carry or store desired program code means in the form of computer-executable instructions or data structures and that can be accessed by a general purpose or special purpose computer.

When information is transferred or provided over a network or another communications connection (either hardwired, wireless, or a combination of hardwired or wireless) to a computer, the computer properly views the connection as a computer-readable medium. Thus, such a connection is also properly termed a computer-readable medium. Combinations of the above should also be included within the scope of

1 computer-readable media. Computer-executable instructions comprise, for example,
2 instructions and data which cause a general purpose computer, special purpose computer,
3 or special purpose processing device to perform a certain function or group of functions.

4 The invention will be described in the general context of computer-executable
5 instructions, such as program modules, being executed by set-top boxes or other
6 computers. Generally, program modules include routines, programs, objects, components,
7 data structures, etc. that perform particular tasks or implement particular abstract data
8 types. The sequence of instructions implemented in a particular data structure or program
9 module represents examples of corresponding acts for implementing the functions or steps
10 described herein.

11 Figure 2 and the corresponding discussion are intended to provide a general
12 description of a suitable environment in which the invention may be implemented. In the
13 discussion, reference is made to a home entertainment system that may be used for
14 displaying and/or recording programming. For purposes of this description and in the
15 claims, a "home entertainment system" may be a display unit, such as a television screen,
16 coupled to a processing device for performing the data processing steps disclosed herein,
17 or may include any number of interconnected consumer electronic devices, one of which
18 having a processing device for performing the data processing steps disclosed herein.
19 Examples of such consumer electronic devices include a video cassette recorder ("VCR"),
20 a video game system, a stereo system, a television or monitor with data processing
21 capabilities, a cable television box, a digital satellite system receiver ("DSS"), a digital
22 video broadcasting system ("DVB"), a digital versatile disc system ("DVD"), a set-top box
23 that serves as an Internet terminal, and any other device capable of processing data as
24 described herein. Furthermore, the term "home entertainment system" is to be understood

as a term that broadly describes a television-viewing environment, whether it is located in a viewer's home, at a place of business, in the public, or at any other location. Also for purposes of this description and in the claims, the term "programming" includes both the viewable portions of moving image data and its associated sound data.

In one embodiment, the present invention is implemented in a system that uses a conventional television screen or other display unit to display information and includes a WebTV® set-top box or a similar Internet terminal that has been adapted to perform the operations that include composing, sending and receiving email, browsing the World Wide Web ("Web"), accessing other segments of the Internet, and otherwise displaying information. An Internet terminal typically uses standard telephone lines, Integrated Services Digital Network (ISDN) lines, cable lines associated with cable television service, or the like to connect to the Internet or other wide area networks.

Figure 2 illustrates a home entertainment system 210 that includes a management system 212, a display device 214 and an audio system 216. Management system 214 may be a set-top box or Internet terminal that has been adapted to perform the operations disclosed herein. Management system 212 may be integrally positioned with or separate from display device 214, which may be a high definition television display, a standard television display, a flat panel display, a projection device, a high definition television display, a computer monitor, or any other device capable of displaying viewable video image data. Audio system 216 may be a speaker, a stereo system, or any device capable of emitting sound data, and similarly may be integrally positioned with or separate from display device 214.

Management system 212 includes a signal input 218, which receives programming from a signal source 220. The programming is transmitted from signal source 220 to

1 signal input 218 via a programming input line 222, which can be a cable or optic
2 connection, a terrestrial antenna system, a satellite system, or any device or system capable
3 of transmitting programming to home management system 212.

4 The signal source 220 may be either a single channel signal source or a multiple
5 channel signal source. A single channel signal source provides programming from a
6 recorded medium, such as a videocassette, compact disc, etc. Examples of a single channel
7 signal source include a VCR, a DVD, and the like. Alternatively, a multiple channel signal
8 source includes any system or device that is capable of sending a signal that may be
9 received by a satellite receiver, a cable or optic connection, a terrestrial antenna, or the
10 like. Examples of a multiple channel signal source include DSS/DVB, a cable box, locally
11 broadcast programming (i.e. programming broadcast using UHF or VHF), and the like.

12 While Figure 2 illustrates home entertainment system 210 as having a single
13 programming input line 222 and a single signal source 220, there can instead be a plurality
14 of programming input lines that transmit programming from a plurality of signal sources.
15 In such embodiments, the home entertainment system may receive the programming from
16 one signal source or from a plurality of signal sources at a time.

17 Management system 212 also includes a user input interface 224, which receives
18 input from an input device 226, such as a remote control, keyboard, microphone, or any
19 other device capable of generating electronic instructions for management system 212.
20 Input device 226 is communicatively coupled to management system 212 over an input
21 link 228 so as to enable such control. Input device 226 generates electronic instructions
22 over input link 228 in response to preprogrammed data or in response to a viewer pressing
23 buttons on input device 226. Input device 226 may also control Web browser software
24 within management system 212 as when management system 212 is a set-top box or an

1 Internet terminal that has been adapted to perform the operations disclosed herein. For
2 instance, input device 226 may be programmed to turn on home entertainment system 210
3 and to tune management system 212 to a channel.

4 Figure 2 illustrates a signal recorder 230, which is capable of receiving video
5 and/or audio data and recording the data on a storage medium. Video signals are
6 transmitted to signal recorder 230 and/or display device 214 by video image link 232,
7 examples of which include a radio-frequency ("RF") link, an S-video link, a composite
8 link, or any other equivalent form of video image link. Similarly, audio link 234 transmits
9 audio data from management system 212 to audio system 216 or to signal recorder 230.

10 The operation of management system 212 is controlled by a central processing unit
11 ("CPU"), illustrated as processing unit 236, which is coupled to an application-specific
12 integrated circuit ("ASIC") 238 and uses computer-executable instructions implemented in
13 software and/or hardwired logic circuitry. Processing unit 236 and ASIC 238 are coupled
14 via a system bus 240, which also interconnects various other system components, such as
15 the system memory 242, mass storage interface 244, user interface 224 and signal input
16 218. Processing unit 236 executes software designed to implement features of
17 management system 212 including features of the present invention. ASIC 238 contains
18 circuitry that is used to implement certain functions of management system 212.
19 Instructions, data, and other software necessary for the operation of processing unit 236
20 and ASIC 238 may be stored in the system memory 242, such as in read-only memory
21 ("ROM") 246 and/or in random-access memory ("RAM") 248, and/or in a mass storage
22 device 250, which is coupled to mass storage interface 244. ROM 246, RAM 248 and
23 mass storage device 250 are communicatively coupled to ASIC 238 so as to be readable by
24

1 ASIC 238 and so that data may be written from ASIC 238 to RAM 248 and possibly mass
2 storage device 250.

3 Mass storage device 250 may be a magnetic hard disk 252 or any other magnetic or
4 optical mass memory device that is capable of storing large amounts of data. Any desired
5 computer-readable instructions or data, including application programs 254, other program
6 modules 256, and an electronic programming guide ("EPG") 258, which specifies the
7 broadcast times and channels of programs can be stored in mass storage device 250. Mass
8 storage device 250 can also be used to record video data 253, in which case, management
9 system 212 generates as a digital video recorder.

10 In the embodiment where management system 212 is associated with the Internet,
11 management system 212 communicates with a remote computer 260 via a wide area
12 network ("WAN") 262 by including a serial port interface 264 that is interposed between
13 the system bus 240 and a modem 266, a wireless link, or other means for establishing
14 communications over a WAN that may be internal or external to management system 212.
15 Management device 212 is also capable of transmitting information via the Internet by
16 direct-dial communication over standard telephone lines, or by using any other available
17 communication medium.

18 While serial port interface 264 may be utilized to connect a modem 266 for
19 communicating across a WAN, serial port interface may also be utilized to connect other
20 consumer electronic devices, such as video game 268, and/or various input devices, such
21 as a keyboard (not shown) or joystick (not shown), to management device 212.

22 Referring now to signal input 218, if the signal on programming input line 222
23 includes multiple channels, a tuner 270 included in signal input 218 tunes to a selected
24 channel in the signal. Multiple tuners 270 can be used to provide enhanced viewing

1 features, such as picture, recording one channel while viewing another, and recording a
2 plurality of channels simultaneously. Video decoder 272 decodes video data from a
3 compressed video format (e.g. MPEG) and performs the subsampling operations described
4 herein. In embodiments where the management system 212 includes multiple tuners 270,
5 management system 212 may also include multiple signal decoders 272 to perform the
6 operations disclosed herein, which enable the image displayed on display device 214 to
7 have a picture-in-picture, split-screen, or other multiple-window format.

8 Management system 212 also includes a video output 274, which may include a
9 video encoder and/or a video converter. The video encoder assigns a code to frames of
10 video data that are transmitted across a video image link 232 and switches between analog
11 and digital formats as necessary. Similarly, audio output 276 can include an audio
12 converter to provide the necessary switching between analog and digital formats.

13 While Figure 2 and the corresponding discussion above provide a general
14 description of a suitable environment in which the invention may be implemented, it will
15 be appreciated that the features of the present invention disclosed herein may be practiced
16 in association with a variety of different system configurations. This is true of the
17 functional components included in Figure 3 as well. The functional components included
18 in Figure 3 illustrate only one example of the practice of the invention. It should be
19 understood that the invention, as embodied in Figure 3, might also be practiced in a
20 configuration that additionally includes one or more of the components as illustrated in
21 Figure 2.

22 As illustrated in Figure 3, the components of an MPEG video bit stream, including
23 frequency coefficients 301, motion vectors 302, and header information 303, are input to
24 decoding system 300. IDCT module 310, subsample module 320, summer 330, and

1 motion vector module 340 are included as components of decoding system 300. IDCT
2 module 310 includes an inverse discrete cosine transform, which converts dequantized
3 frequency coefficients 301 into spatial samples 306. Motion vector module 340 uses
4 header information 303 and motion vectors 302 to retrieve fetched prediction data 305,
5 which is information from previously decoded images. Subsample module 320
6 subsamples spatial samples 306 by a selected subsampling factor. Summer 330 sums
7 subsampled spatial samples 307 and fetched prediction data 305 to generate reconstructed
8 output 304, which includes a sequence of video images that have a size smaller than the
9 size of the corresponding video images that were originally encoded. The sequence of
10 video images can then be displayed using an NTSC or other standard definition television
11 ("SDTV") display device, using picture-in-picture, or otherwise processed or displayed as
12 desired.

13 IDCT module 310, subsample module 320, summer 330 and motion vector module
14 340 may be software components, hardware components, or a combination of software and
15 hardware. Although these modules are shown as separate modules, one or more of their
16 functions may be combined in the components of Figure 2, including signal input 218,
17 processing unit 236, system memory 242, tuner(s) 270, signal decoder 272, video output
18 274, audio output 276, ASIC 238 or any combination thereof so as to be inseparably
19 integrated. These modules may also be stored on mass storage device 250 and included in
20 applications programs 254 or other program modules 256.

21 In operation, IDCT module 310 receives frequency coefficients 301 in parallel with
22 motion vector module 340 receiving motion vectors 302 and header information 303.
23 IDCT module 310 converts frequency coefficients 301 into spatial samples 306. While an
24 inverse discrete cosine transform is frequently used to convert between the frequency and

1 spatial domains, practice of the invention is not limited to this method of conversion. The
2 invention may be practiced using any component that converts a frequency domain
3 representation of data to a spatial domain representation of the same data, including
4 methods using Fourier transforms.

5 Subsample module 320 subsamples spatial samples 306 to reduce the amount of
6 data. Subsampling may be performed in any number of ways. One method is to
7 subsample in only one direction, either horizontally or vertically. Subsampling in one
8 direction results in a reduction in the data sent to summer 330 equal to the inverse of the
9 subsampling rate. For example, subsampling by a factor of two, in the horizontal or
10 vertical direction, results in sampling every other value in spatial samples 306 in the
11 selected direction, which reduces the data sent to summer 330 to one half its value prior to
12 being processed by subsample module 320.

13 Subsampling may also be performed in both the horizontal and vertical directions
14 simultaneously. Subsampling in two directions results in a reduction in the data sent to
15 summer 330 equal to the inverse of the product of the subsampling rate in the horizontal
16 direction times the subsampling rate in the vertical direction. Subsampling by a factor of
17 two in the horizontal and vertical directions simultaneously results in sampling every other
18 value in spatial samples 306 in both the horizontal and vertical directions. This reduces the
19 data sent to summer 330 to one fourth its value prior to being processed by subsample
20 module 320.

21 Figures 5a-5c illustrate various techniques for subsampling spatial samples and
22 applying appropriate weighting factors to the spatial samples according to selected
23 embodiments of the invention. It is noted that the following subsampling techniques
24 represent only a few of the large number of subsampling algorithms and processes that can

be used with the invention to decimate video data. One basic principle illustrated by the following subsampling techniques is that spatial samples having been generated by the IDCT are selected and weighted to generate the subsampled spatial samples.

Figure 5a illustrates a process for decimating spatial samples in the horizontal direction by a factor of two. Spatial samples A-D 350a-d represent blocks in a macroblock of a frame of MPEG video data and have a color parameter, either a luminance or chrominance. Spatial samples A-D 350a-d are subsampled by a factor of two to generate subsample A' 352 and subsample B' 354. The color parameters specified by sample A 350a and sample B 350b are combined in order to generate one subsample A' 352 from the two spatial samples. In order to preserve the spatial positioning of sample A 350a and sample B 350b, the color parameters are weighted by a factor inversely proportional to the relative spatial distance from the samples to the subsample in this example. In this case, the relative spatial distances of sample A 350a and sample B 350b to subsample A' 352 are equal. Accordingly, the color parameters are multiplied by the same factor, namely, $\frac{1}{2}$. Thus, the color parameter of subsample A' 352 is generated using the equation $A' = \frac{1}{2} A + \frac{1}{2} B$, where A' represents the color parameter of subsample A' 352 and A and B represent the color parameters of sample A 350a and sample B 350b, respectively. Similarly, the color parameter of subsample B' 354 is generated using the equation $B' = \frac{1}{2} C + \frac{1}{2} D$.

Figure 5b illustrates a process for decimating spatial samples of non-interlaced video data in the vertical direction by a factor of two. Spatial samples A-D 360a-d represent spatial samples in successive rows of the non-interlaced video data. The process of obtaining the color parameters of subsample A' 362 from sample A 360a and sample B 360b and of subsample B' 364 from sample C 360c and sample D 360d is similar to the horizontal decimation process described above in reference to Figure 5b.

Figure 5c illustrates a process of decimating spatial samples of field-interlaced video data in the vertical direction by a factor of two. Spatial sample A 370a and spatial sample C 370c represent samples in successive rows of a first field of the frame of video data, while spatial sample B 370b and spatial sample D 370d represent samples in successive rows of a second field. In this example, the samples of the successive rows are used to generate the corresponding subsample. In order to preserve the spatial positioning of sample A 370a and sample C 370c, the color parameters are weighted by a factor inversely proportional to the relative spatial distance from the samples to the subsample in this example. In this case, the relative spatial distance of sample A 370a to subsample A' 372 with respect to the spatial distance of sample C 370c to subsample A' 372 is 1:3. Accordingly, the color parameter of sample A 370a is multiplied by a factor that is inversely proportional to the ratio 1:3, namely, $\frac{3}{4}$. Similarly, the color parameter of sample C 370c is multiplied by a factor that is inversely proportional to its relative distance to subsample A' 372, namely, $\frac{1}{4}$. Thus, the color parameter of subsample A' 372 is generated using the equation $A' = \frac{3}{4} A + \frac{1}{4} C$, where A' represents the color parameter of subsample A' 372 and A and C represent the color parameters of sample A 370a and sample C 370c, respectively. Similarly, the color parameter of subsample B' 374 is generated using the equation $B' = \frac{1}{4} B + \frac{3}{4} D$.

The foregoing techniques can be combined to decimate in both the horizontal and vertical directions. Although the foregoing subsampling techniques can be used advantageously with the invention, the specific algorithms for subsampling are not critical to the invention. Moreover, other weighting factors can be used to filter the spatial samples as desired.

1 Motion vector module 340 combines data from motion vectors 302 and header
2 information 303 to generate fetched prediction data 305. Data stored in motion vectors
3 302 is used to modify header information 303 to generate predictive and bi-directionally
4 predictive frames. Motion vector module 340 decimates, scales, or otherwise adjusts
5 motion vectors 302 by the same factor as the subsampling factor used by subsample
6 module 320, so that the adjusted motion vectors can be used to fetch, from a reference
7 frame, prediction data 305 that corresponds with the subsampled spatial samples 307. For
8 instance, if subsample module 320 subsamples in the horizontal direction by a factor of
9 two, motion vector module 340 decimates the motion vectors in the horizontal direction by
10 a factor of two. Similarly, if subsample module 320 subsamples in both the horizontal and
11 vertical directions, motion vector module 340 decimates motion vectors 302 by the same
12 factors in the horizontal and vertical directions. Summer 330 sums subsampled spatial
13 samples 307 and fetched prediction data 305 to form reconstructed output 304.

14 The operation of the structure of Figure 3 will now be described with respect to
15 Figure 4, which is a flowchart of a method for subsampling and decoding an MPEG video
16 bit stream. The method of this embodiment of the invention begins by performing, in
17 parallel, an act of processing motion vectors in order to fetch prediction data from a
18 reference frame (act 401) and an act of processing frequency coefficients of the MPEG
19 video bit stream with an IDCT to generate IDCT output data (act 403). The method also
20 includes an act of processing a reference frame and the decimated or otherwise adjusted
21 motion vectors in a frame prediction module to generate predicted frame data (act 402) and
22 an act of decimating the IDCT output data (act 404). The decimated IDCT output data and
23 the predicted subsampled frame data are then summed to generate video images having
24 reduced size (act 405).

By decimating the video data in the foregoing manner, an MPEG video bit stream having an HDTV format can be subsampled and decoded so as to obtain a sequence of video images having sizes such that the images can be rendered on an SDTV display device or in a picture-in-picture window. Subsampling, decoding and displaying the video data in this manner eliminates the need to decode and buffer the full volume of the original HDTV video data. For instance, if an HDTV MPEG video bit stream is received by the device that decodes the video data (e.g., management system 212) at a rate of 90MB/sec, the processing capabilities of the decoder and the associated memory can be significantly smaller than that which has been required using prior art techniques. For instance, if the video data is subsampled by a factor of two in both the horizontal and vertical directions, the frame buffers used to temporarily store the frames of the video data prior to display on the display device can have a total storage capacity as small as 1.6 MB or smaller. Similarly, the internal data transmission rates at which the decoded video data is transmitted can be reduced by a factor of four using the two-dimensional decimation.

Because subsampling is performed after the frequency coefficients are processed by IDCT module 310, that is by subsampling spatial samples 306, the resolution of the decoded image data is typically reduced. However, because one useful implementation of the invention is to process video data for display on a display device or a region thereof having a resolution that is often significantly smaller than the resolution of the full volume of video data, this loss of image quality is often negligible. In addition, unlike many other decimation processes that may be technically feasible or which are known in the art, subsampling performed in the spatial domain does not result in a significant compounded loss of image quality for relatively static or constant pans.

1 The present invention may be embodied in other specific forms without departing
2 from its spirit or essential characteristics. The described embodiments are to be considered
3 in all respects only as illustrative and not restrictive. The scope of the invention is,
4 therefore, indicated by the appended claims rather than by the foregoing description. All
5 changes, which come within the meaning and range of equivalency of the claims, are to be
6 embraced within their scope.

7 What is claimed and desired secured by United States Letters Patent is:
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

WORKMAN, NYDEGGER & SEELEY
A PROFESSIONAL CORPORATION
ATTORNEYS AT LAW
1000 EAGLE GATE TOWER
60 EAST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84111